
UnIT-NET IIDE : Infrastructure nationale ukrainienne pour l'intraéchange de données électroniques

Ermolayev Vadim *, Spivakovsky Alexandre **, Zholtkevych Grygoriy ***

*Université d'Etat de Zaporojié . Département de Technologie d'Information
66, rue Zhukovsky. 69063 Zaporojié UKRAINE

** Université d'Etat de Kherson
27, rue 40 years Octobre 73000 Kherson UKRAINE

***Université Nationale Karazin de Kharkiv
4,pl. Svobody 61077 Kharkiv UKRAINE

eva@zsu.zp.ua ; spivakovsky@ksu.ks.ua ; Gregory.N.Zholtkevych@univer.kharkov.ua

Sections de rattachement : 27, 61, 6
Secteur : Secondaire

RÉSUMÉ. L'intraéchange de données électroniques (IDE) est ici l'échange d'ordinateur-à-ordinateur de données entre les universités et les organismes d'état Ukrainiens. Nous présentons l'architecture de référence d'IIDE qui est l'infrastructure de logiciel soutenant IDE, c'est-à-dire le système d'information distribuée à plusieurs couches comportant les serveurs de logiciel, les services, les composants et les outils pour fournir la recherche intelligente et ontologiquement-conduite de l'information des Ressources d'Information (RI) distribuées, hétérogènes, légalement et physiquement autonomes dans le cadre du Système National d'Enseignement Supérieur. Les principes de la conception architecturale, les processus qui définissent la fonctionnalité d'IIDE, la hiérarchie architecturale, les composants, les ontologies d'IIDE y sont présentés. IIDE servira de noyau pour fournir les services d'intraéchange des données pour la gestion de l'Université et des groupes de recherche distribués travaillant sur des projets communs.

MOTS-CLÉS : UnIT-Net, IIDE, l'architecture, le médiateur, l'emballage, le système d'information, l'ontologie.

1. Introduction

Pour atteindre et soutenir le perfectionnement dynamique, les organismes service-orientés tels que les universités, ont besoin d'une infrastructure qui soutienne la gestion flexible et robuste de leurs activités et soit un support à la prise de décision. Les activités des universités ainsi que leur coordination et le contrôle au niveau national

comportent le traitement des données et de la connaissance d'entreprise. Les organismes impliqués dans le cadre éducatif étant légitimement indépendants, ils possèdent et maintiennent leurs sources de données et de connaissance de façon autonome ; c.-à-d. indépendamment les uns des autres et, à un degré élevé, du corps de coordination, tel que le Ministère National. Le fait que ces ressources d'information sont autonomes implique des complications sérieuses pour leur intégration. IIDE est l'infrastructure de logiciel prévoyant l'échange de données électroniques entre les universités et les organismes d'état d'Ukraine. Plus précisément, IIDE est le système d'information distribuée à plusieurs niveaux comportant les serveurs de logiciel, les services, les composants et les instruments fournissant la récupération intellectuelle et ontologiquement conduite de l'information des Ressources d'Information (RI) distribuées, hétérogènes, légalement et physiquement autonomes dans le cadre d'organisation du Système National de l'Enseignement Supérieur.

Le reste de l'article est structuré d'après le plan ci-dessous. La section 2 décrit le travail relatif et les principes utilisés dans la conception architecturale d'IIDE. La section 3 esquisse l'architecture de UnIT-NET IIDE. La section 4 se concentre sur la famille des ontologies qui conduisent la décomposition de question, questionnent la recherche documentaire de traduction et les résultats marquant dans IEDI. La section 5 donne des remarques de conclusion et décrit les directions du futur travail.

2. Le travail relatif

Dans le contexte décrit le genre de IIDE peut être classifié comme le domaine intelligent distribué de la recherche d'information (I2R) dans le domaine plus large de l'intégration intelligente de l'information (I3). Les activités de recherche dans ce domaine ont été intenses dans la décennie passée, particulièrement dans la ligne de l'action principale de « Information Society Technologies » de FP6 de l'UE et des cadres nationaux et internationaux semblables. Les exemples des projets R&D développant les cadres formels, algorithmiques, architecturaux, déployant des prototypes de logiciel pour I2R des RI distribuées et hétérogènes et l'intégration intelligente de l'information (I3) sont BUSTER (Stuckenschmidt et al., 2000), DOME ((Cui et al., 2001), (Cui et al., 2002)), InfoSleuth (Bayardo et al., 1997), MOMIS (Bergamaschi et al., 1998), OBSERVER (Kashyap et al., 2000), Ontobroker (Decker et al., 1999), PICSEL (Lattes et al., 2000), SIMS (Arens et al., 1996), et d'autres. Un bon aperçu des approches ontologiquement basées à I2R et à I3 peut être trouvé en (Wache et al., 2001). UnIT-Net IIDE construit son architecture sur les leçons apprises de ces projets. Cependant le projet RACING a exercé l'influence principale sur l'architecture d'IIDE. Dans le projet RACING une architecture décentralisée de médiateur prévoit pour chaque ressource d'information un agent/emballage séparé, qui stocke des affichages entre les ontologies globales/communes et la ressource d'information située au-dessous (Ermolayev et al., 2003). Certains problèmes mentionnés dans les sources citées ci-dessus sont pour le moment partiellement résolus. Ces solutions partielles

limitent évidemment le domaine d'application et la fonctionnalité des prototypes de logiciel déployés pour I2R. Les limitations pour IIDE sont nommées ci-dessous: IIDE est construit selon les principes d'architecture de médiateur-emballage (Wiederhold, 1992) avec le médiateur centralisé; IEDI exploite l'approche hybride (Wache et al., 2001) pour la représentation de la connaissance; IEDI emploie l'enregistrement de ressource d'information pour permettre à la ressource de devenir disponible pour la requête; IEDI ne fournit pas la pleine automatisation pour l'affichage et l'alignement des ontologies; les composants d'IIDE emploient des techniques réinscriptibles avec des affichages pour produire, traiter, et exécuter des requêtes. La tâche fut d'élaborer le prototype de logiciel pour montrer la praticabilité de l'approche ontologiquement conduite à I2R et, puis ensuite en IDE entre les universités et les organismes d'état au niveau national.

3. Architecture de référence de IIDE

Le but principal d'IEDI est de prévoir l'exécution de requêtes de composition de RI de pré-enregistrés, mais indépendants, distribuées et sémantiquement hétérogènes. Cela implique que IIDE est naturellement un système réparti. La requête peut exiger de renouveler les données de plusieurs RI géographiquement distribuées, qui appartiennent à différents propriétaires juridiques et sont physiquement stockés en divers lieux. C'est pourquoi les processus de IIDE se composent d'un certain nombre de tâches et d'activités exécutées aux noeuds distribués. Ces tâches devraient certainement être exécutées d'une manière contrôlée et (commandée) ordonnée. Un processus implique normalement des activités automatisées exécutées par le logiciel d'IIDE et des activités humaines, comme la fusion d'ontologie et l'alignement, fournis avec des méthodologies appropriées et des outils de logiciel. Des activités humaines sont exécutées par différents utilisateurs : utilisateur autorisé (UA), ingénieur d'ontologie de médiateur (IOM), ingénieur d'ontologie de RI (IORI), fournisseur de IR (Ermolayev et al. 2004).

Le facteur important qui a sérieusement influencé la création de l'architecture de IIDE est l'hétérogénéité sémantique des RI qui sont enregistrées au médiateur d'IEDI. Cela a impliqué l'utilisation de la hiérarchie des ontologies qui dirigent réellement l'exécution des requêtes distribuées aux différentes RI. Les tâches de fusionnement et d'alignement des ontologies décrivant la sémantique des RI et l'ontologie commune du médiateur – Ontologie du domaine du médiateur (ODM) - sont exécutés manuellement. IIDE fournit des ontologies et des outils de référence pour des activités de cette technologie d'ontologie. Malgré tout restent à exécuter : la formulation de requête, l'extraction de sous-requête, des tâches d'exécution de sous-requête d'une façon franche et presque automatique. Le diagramme du scénario d'exécution de requête de IIDE est donné sur le figure 1. Les diagrammes pour l'Enregistrement des RI et l'Entretien de concordance d'ontologie sont omis ici faute d'espace (Ermolayev et al. 2004).

La hiérarchie architecturale d'IIDE est définie selon l'analyse des processus et des tâches d'IIDE et reflète le type médiateur-emballage d'architecture d'IIDE. La hiérarchie

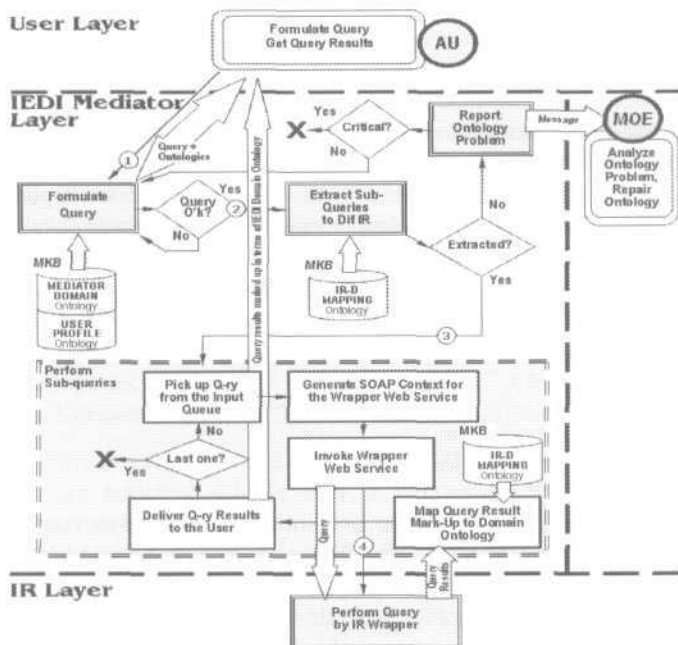


Figure. 1. Diagramme de processus pour le scénario d'exécution de demande d'IIDE.

représente l'organisation globale d'IIDE, elle est décrite ci-dessous : Quels sont les Composants, les Outils et les rôles d'utilisateur aux couches spécifiques d'IIDE ? Comment les Clients et les Serveurs d'IIDE interagissent-ils à travers les couches de leur architecture ?

La couche d'utilisateur d'IIDE est l'environnement pour UA et ses clients. L'emballage de RI d'IIDE et les couches de RI représentent les supports autonomes, hétérogènes, et distribués de RI.

La couche de médiateur d'IEDI est le support pour les composants et les outils fournissant les moyens de médiation entre les US formulant les commandes et recherchant les résultats des RI enregistrées et des emballages respectifs des RI pour fournir les informations appropriées. La figure 2 représente la structure d'IIDE.

Les composants de logiciel d'IIDE sont divisés en deux catégories : Clients et Serveurs selon leur fonctionnalité. Les clients d'IIDE sont liés aux UA d'IIDE et fournissent les interfaces pour leurs activités. Le client d'UA fournit des interfaces d'IIDE pour un UA. Il fonctionne dans l'environnement générique de web browser (+ Java Virtual Machine) à la couche d'utilisateur d'architecture d'IIDE (figure 2) et fournit les interfaces pour les tâches de : Formulation de requête d'Utilisateur, Approbation de requête d'Utilisateur, visionnage des résultats de requête. Le client d'UA interagit avec l'outil de formulation de requête d'IIDE et avec les composants suivants d'IEDE :

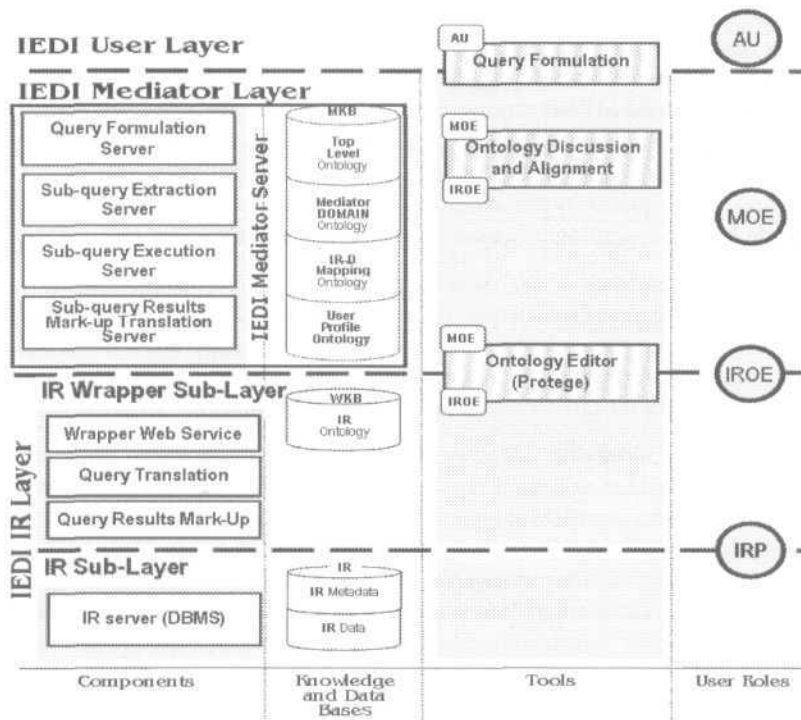


Figure 2. Clients et serveurs d'IIDE le long des couches de l'architecture.

Serveur d'accès de médiateur d'IIDE et Serveur de formulation de requête (le composant du serveur de médiateur d'IIDEI). Le client d'IOM fournit des interfaces d'IIDE pour l'IOM. Il fonctionne dans l'environnement de Java Virtual Machine (JVM) à la couche de médiateur d'architecture d'IIDE et fournit les interfaces pour les tâches de la discussion, de la fusion, de l'alignement, de l'édition et de la réparation d'Ontologies d'IIDI. Le client d'IORI fournit les interfaces d'IIDE pour un IORI et est semblable à un client d'IOM. Il fonctionne aux couches de médiateur et d'emballage de RI d'architecture d'IIDE et fournit les interfaces pour les tâches de la discussion, de l'édition et de la réparation ainsi que pour la négociation sur ORI - fusion de ODM dans le processus d'enregistrement de RI. Les clients de IOM et de IORI interagissent avec les outils suivants de IIDE : Discussion d'Ontologie et alignement (en cours de développement), Rédacteur d'Ontologie (Protégé (Fridman-Noy et al., 2001)). Les clients de IOM et de IORI interagissent avec les composants suivants de IIDE: Serveur d'Accès de Médiateur d'IIDE. Les clients et les serveurs d'IIDE sont énumérés figure 2.

4. Ontologies d'IIDE

IIDE par son rôle est le système réparti de médiateur fournissant un certain genre d'intégration sémantique d'information recherchée des ressources distribuées,

hétérogènes, et autonomes d'information. C'est pourquoi l'exécution et l'utilisation propre des descriptions sémantiques de cette information est le problème critique pour l'accomplissement du système d'IIDE. On suppose que des descriptions sémantiques en IIDE sont formalisées et maintenues comme ontologies de OWL (OWL, 2003) à différentes couches de l'architecture. L'architecture d'IIDE emploie l'approche hybride (Wache et al., 2001) pour la description explicite de la sémantique de ressource d'information. Les quatre types d'ontologies sont fournis : ontologie supérieure, ontologie de domaine, ontologie de ressource et ontologie de référence.

L'ontologie supérieure (OS) définit les éléments supérieurs basiques. Ils sont employés dans le processus d'affichage des éléments d'ontologie de ressource à ceux de domaine. L'OS sert de base à la discussion sur chaque concept entre IOM et IORI. Elle permet à deux ontologies quelconques d'IIDE d'être comparables. La conception d'ontologie supérieure d'IIDE est basée sur DOLCE (Masolo et al., 2002).

L'ontologie de domaine (MOD) représente la connaissance particulière de domaine. Il y a plusieurs raisons d'explorer l'ontologie de domaine dans le médiateur d'UNIT-NET IIDE. La première est que la MOD fournit aux UA l'occasion de formuler leurs requêtes en utilisant des concepts, convenus au sein de la communauté de domaine, et de stocker des correspondances entre la connaissance personnelle d'utilisateur sur le domaine et l'ontologie convenue de domaine dans leurs profils d'utilisateur (l'ontologie de référence de profils d'utilisateur – ORPU). Une autre raison est que l'ontologie de domaine présente une vision de la communauté sur le domaine, et joue donc un rôle éducatif.

L'ontologie de ressource d'information (ORI) est un genre d'ontologie de domaine, qui est construit sur le côté de ressource indépendamment d'autres ressources aussi bien que des ontologies de médiateur. Il présente la vision d'IOI sur le domaine. L'ORI est employé en cours d'enregistrement de ressource au médiateur. Chaque ressource d'information enregistrée devrait avoir sa propre ontologie de ressource.

Ontologies de référence (IR – Ontologie de l'affichage de domaine (IROAD), ORPU) sont des ontologies de médiateur, qui stockent la connaissance sur des correspondances entre les concepts dans deux ontologies ou plus. IROAD contient des axiomes sur équivalence/dérivation entre concepts/slots. La fonction de l'ORPU est de représenter la sémantique des profils d'UA (Ermolayev et al. 2003).

Le tableau 1 récapitule la participation des ontologies de médiateur d'IIDE dans des processus d'IEDI.

5. Conclusion

Dans cet article a été présentée l'architecture de référence d'IIDE qui est le système de logiciel distribué à plusieurs couches comportant les serveurs de logiciel, les services, les composants et les outils pour fournir la récupération intellectuelle et

Tableau 1.

| <i>Processus</i> \ <i>Ontologies</i> | <i>TLO</i> | <i>MO DCo re</i> | <i>MO D</i> | <i>IROAD</i> | <i>ORI</i> | <i>ORP U</i> |
|--|------------|--------------------------|-----------------|--------------|------------|------------------|
| <i>Requête les RI distribués, sémantiquement hétérogènes et autonome</i> | -- | R | R | R | R | R/U |
| <i>Enregistre la nouvelle ressource d'information</i> | R | R | R/U | R/U | R | -- |
| <i>Maintient la cohérence dans des descriptions sémantiques</i> | R | R/U | R/U | R/U | R/U | R/U |

Légende : R – utilisé pour la référence seulement, R/U – utilisé comme une référence et est renouvelé, -- – non utilisé.

ontologiquement-conduite de l'information des RI distribuées, hétérogènes, légalement et physiquement autonomes pour le réseau organisationnel du Système National d'Enseignement Supérieur dans le cadre du projet UnIT-Net. Cette architecture est construite sur les principes ci-dessous : c'est l'architecture de médiateur-emballage avec le médiateur centralisé; elle exploite l'approche hybride pour la représentation de la connaissance; elle emploie l'enregistrement de l'information de ressource pour permettre à la ressource de devenir disponible pour la requête; IIDE combine les processus exécutés automatiquement (traitement de requête distribuée ontologiquement conduite) ainsi que manuellement (discussion d'ontologie, fusionnement, affichage, et alignement pendant l'enregistrement de RI et le maintien de cohérence d'ontologie); ses composants utilisent des techniques réinscriptibles d'affichage pour produire, traiter, et exécuter des requêtes.

Du point de vue sémantique IIDE exploite la hiérarchie des ontologies qui sont remplies incrémentalement tandis que les nouvelles RI sont enregistrés au médiateur. Ces ontologies sont employées à tour de rôle pour diriger la fonction principale du médiateur : assister à la formulation de requête, pour décomposer la requête en ensemble de sous-requêtes (une par RI approprié), pour transporter les sous- requêtes aux emballages de RI respectifs, pour traduire la sous- requête au niveau d'emballage de RI, pour marquer des résultats de requête.

Les plans à court terme pour les futurs travaux comportent l'accomplissement du prototype de recherches d'IIDE et son évaluation à la collection initiale de RI des universités membres du Consortium UnIT-Net.

Remerciements

Ce travail a été financé par European Training Foundation dans le cadre du projet UnIT-Net. Les auteurs voudraient remercier toutes les universités dont le travail a contribué aux Spécifications d'Architecture de Référence d'IIDE (Ermolayev et al. 2004) ainsi que Kseniya Ermolayeva qui a traduit cet article en Français.

Bibliographie

- Arens, Y. et al.: Query Reformulation for Dynamic Information Integration. *Journal of Intelligent Information Systems*. 1996.
- Bayardo et al.: *InfoSleuth: Semantic Integration of Information in Open and Dynamic Environment*. In Proceedings of the 1997 ACM International Conference on the Management of Data (SIGMOD), Tucson, Arizona, May 1997.
- Bergamaschi, S. et al.: An Intelligent Approach to Information Integration. In: Proc. Of Formal Ontology in Information Systems (FOIS-98), June, 1998.
- Cui, Z. et al.: Issues in Ontology-based Information Integration. In: (A. Gomez-Perez, M. Gruninger, H. et al.) Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, USA, August 4-5, 2001, p. 141-146.
- Cui, Z. et al.: Semantic B2B Integration: Issues in Ontology-based Applications. *SIGMOD Record*, Vol.31, No.1, March 2002, p. 43-48.
- Decker, S. et al.: Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. In R. Meersman et al. (eds.): *Semantic Issues in Multimedia Systems*. Proceedings of DS-8. Kluwer Academic Publisher, Boston, 1999, p. 351-369.
- Ermolayev, V. et al. Capturing Semantics from Search Phrases: Incremental User Personification and Ontology-Driven Query Transformation. In: Proc. of the 2-nd Int. Conf. on Information Systems Technology and its Applications (*ISTA'2003*), Kharkiv, Ukraine, June 19-21, 2003, p. 9-20, LNI Series, ISBN 3-88579-359-8.
- Ermolayev, V., et al.: The Infrastructure for Electronic Data Interchange. Reference Architecture Specification. Version 1.0. UNIT-NET Deliverable No D2.2.D.1. 2004, URL: <http://www.compscipreprints.com/comp/Preprint/eva/20040228/1>.
- Kashyap V., Sheth A.: *Information Brokering across Heterogeneous Digital Data: A Metadata-based Approach*. Kluwer Academic Publishers, 2000.
- Lattes V., Rousset M.-C.: The Use of CARIN Language and Algorithms for Information Integration: The PICSEL System. *International Journal of Cooperative Information Systems*, Vol.9, No.4, 2000, p. 383-401.
- Masolo, C., et al.: WonderWeb Project Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology.
- Fridman-Noy, N. et al.: Creating Semantic Web Contents with Protege-2000. *IEEE Intelligent Systems* 16(2), p. 60-71, 2001.
- OWL Web Ontology Language Reference. W3C Proposed Recommendation 15 December 2003. URL: <http://www.w3.org/TR/owl-ref/>.
- Stuckenschmidt H. et al.: Enabling technologies for interoperability. In: (Visser, U., Pundt H. Eds.) Workshop on the 14th International Symposium of Computer Science for Environmental Protection, Bonn, Germany, 2000, p. 35-46.
- Wache, H. et al. : Ontology-Based Integration of Information - A Survey of Existing Approaches. In: (A. Gomez-Perez et al.) Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, USA, August 4-5, 2001, p. 108-118.

Wiederhold, G.: Mediators in the Architecture of Future Information Systems. *IEEE Computer*, 25, 3 (March), 1992, p. 38–49.